Experiences with Aber-OWL, an Ontology Repository with OWL EL Reasoning

Slater, Luke; Rodríguez-García, Miguel Ángel; O'Shea, Keiron; Schofield, Paul N.; Gkoutos, Georgios V.; Hoehndorf, Robert

DOI:
10.1007/978-3-319-33245-1_8

Experiences with Aber-OWL, an ontology repository with OWL EL reasoning

Luke Slater[1], Miguel Angel Rodriguez Garcia[1], Keiron O'Shea[1,2], Paul N Schofield[3], Georgios V Gkoutos[2], and Robert Hoehndorf[1]

[1] Computational Bioscience Research Center, King Abdullah University of Science and Technology, Thuwal 23955-6900, KSA
{luke.slater,miguel.rodriguezgarcia,robert.hoehndorf}@kaust.edu.sa
[2] Aberystwyth University, Aberystwyth, SY23 3DB, Wales, UK,
keo7@aber.ac.uk
[3] University of Cambridge, Downing Street, CB2 3EG, England, UK
pns12@hermes.cam.ac.uk

Abstract. Ontologies are widely used in biology and biomedicine for the annotation and integration of data, and hundreds of ontologies have been developed for this purpose. These ontologies also constitute large volumes of formalized domain knowledge, usually expressed in the Web Ontology Language (OWL). Computational access to the knowledge contained within them relies on the use of automated reasoning. We have developed Aber-OWL, an ontology repository that provides OWL EL reasoning to answer queries and verify the consistency of ontologies. Aber-OWL also provides a set of web services which provide ontology-based access to scientific literature in Pubmed and Pubmed Central, SPARQL query expansion to retrieve linked data, and integration with Bio2RDF. Here, we report on our experiences with Aber-OWL and outline a roadmap for future development.

Keywords: biomedical ontology, Semantic Web, literature search, semantic indexing, query expansion

Introduction

Ontologies are used in most biological databases for the annotation and integration of data, and hundreds of ontologies have been developed for that purpose. These ontologies are commonly expressed in either the Web Ontology Language (OWL) [6] or an OWL-compatible language such as the OBO Flatfile Format [12]. Ontology repositories, such as BioPortal [14], the Ontology Lookup Service (OLS) [5] and OntoBee [18], currently provide web services and interfaces to access ontologies and their data in the biological domain. However, they do not utilize reasoning in the services they provide, and thus do not provide the advantages of semantic access, access to inferred knowledge and consistency verification.

To enable this, we have created Aber-OWL[10] – an ontology repository in which access to ontologies is underpinned by reasoning. Aber-OWL consists

primarily of an API, a web repository and a set of web services that provide ontology-based access to biological and biomedical data and literature. Here, we discuss our experiences with developing an ontology portal based on automated reasoning, discuss the current limitations, and suggest future extensions.

An overview of Aber-OWL

Reasoning services

The main component of Aber-OWL is a server that provides access to a large set of ontologies (currently 391) through an OWL EL reasoner. Ontologies are classified at the beginning of the server's runtime, and then kept in memory. We use the ELK reasoner [13], which supports the OWL EL profile, and any axioms that do not fall within the OWL EL subset are ignored. The restriction to OWL EL expressivity ensures that classification and query times remain tractable.

Access to the classified ontologies is provided through a REST API. This API can be utilized to perform Description Logic queries; specifically, it can be used to retrieve sub-, super-, or equivalent classes of a class description (which must also fall in the OWL EL profile). Querying is performed by transforming a class description in Manchester OWL Syntax [11] into an OWL class expression using the OWL API. If this transformation fails (e.g., when the query string provided is not a valid OWL class expression within the ontology being queried), an empty set of results is returned.

If the transformation succeeds, the ELK reasoner is used to retrieve sub-, super- or equivalent classes of the OWL class expression. Each query can be performed over a single or multiple ontologies stored within Aber-OWL. Consequently, results may be returned from multiple different ontologies at once. If a URL is specified as part of a query but the ontology is not available within Aber-OWL's repository, an attempt is made to retrieve the ontology from the URL, classify the ontology, perform the query over this ontology and return the results automatically. The API also provides additional ways to access the content of the ontologies, such as a substring-based search for classes, retrieving class descriptions based on the class IRI, and others.

Ontology-based data access

One of our main aims in developing Aber-OWL is to demonstrate the potential for ontology-based access [4] to biological and biomedical data. Therefore, we developed several webservices that make use of Aber-OWL and combine OWL EL reasoning with access to different types of data sources.

The Aber-OWL: PubMed service is built on top of the Aber-OWL reasoning infrastructure, and retrieves articles in PubMed and PubMed Central in which any of the labels and synonyms of classes returned by a given semantic query appear. The literature search is performed over an Apache Lucene index holding all full text articles in PubMed Central and all abstracts in PubMed (using a

disjunctive Lucene query of the class labels in the result set of the Aber-OWL query). This service allows, for example, to retrieve all articles that mention a subclass of `part-of some Heart` in its text.

The Aber-OWL: SPARQL service performs query expansion on a SPARQL query to incorporate the results of an Aber-OWL query. In particular, the set of class IRIs returned by an Aber-OWL query can be bound to a variable in SPARQL (using the SPARQL 1.1 `VALUES` statement) or used as an RDF collection that could, for example, be used with the `IN` operator as part of a `FILTER` statement. The use of Aber-OWL: SPARQL allows, for example, to query the UniProt [16] SPARQL endpoint for all proteins that have as their function a part of apoptosis that also regulates apoptosis (`part-of some 'apoptotic process' and regulates some 'apoptotic process'`). We further incorporated direct access to Bio2RDF [3] based on either the IRI of a class returned by an Aber-OWL query or based on the label of the class.

Experiences

One of the main challenges in developing an ontology portal based on OWL reasoning is the usability. Our target audience for Aber-OWL is twofold: on one side, we aim to provide services to bioinformaticians and ontologists who wish to make use of automated reasoning over ontologies as part of their workflow, and on the other hand, we aim to provide a useful repository of ontologies for biologists and biomedical researchers. While the first group of users will primarily use the API provided by Aber-OWL, the second group would rely mainly on the user interfaces we provide. However, making Description Logic querying easily accessible to a wide range of users through a common user interface is challenging and has constituted the main criticism we have received so far. To address these challenges in the future, we are considering utilizing natural language query interfaces [17], or visual construction of DL queries.

A related challenge is the automatic identification of labels and descriptions in ontologies. Across the range of over 390 ontologies in Aber-OWL, several different annotation properties are used to characterize the labels and textual descriptions of classes and object properties. Since Manchester OWL syntax relies on identifying natural language labels for classes so that they can be used as part of a class description, it is crucial to find a unified way of identifying labels, synonyms and descriptions of classes. The annotation properties we currently use to identify these are shown in Table 1, and they cover most of the ontologies in Aber-OWL. With the broad range of ontologies in Aber-OWL, the annotation properties in use will have to be constantly updated. As an intermediate solution, we now allow queries to be submitted in two forms, using either the labels of the classes and object properties, or using their IRIs directly.

With a more widespread adoption of Aber-OWL, we also have the potential for collecting a large set of real world Description Logic queries together with their execution time, which may become a useful resource for Description Logic reasoner performance evaluation [2]. We have created a log of all

Table 1. Labels, Synonyms and Descriptions used in Aber-OWL

Labels:
`rdfs:label`
http://www.w3.org/2004/02/skos/core#prefLabel
http://purl.obolibrary.org/obo/IAO_0000111
Synonyms:
http://www.w3.org/2004/02/skos/core#altLabel
http://purl.obolibrary.org/obo/IAO_0000118
http://www.geneontology.org/formats/oboInOwl#hasExactSynonym
http://www.geneontology.org/formats/oboInOwl#hasSynonym
http://www.geneontology.org/formats/oboInOwl#hasNarrowSynonym
http://www.geneontology.org/formats/oboInOwl#hasBroadSynonym
Descriptions:
http://purl.obolibrary.org/obo/IAO_0000115
http://www.w3.org/2004/02/skos/core#definition
http://purl.org/dc/elements/1.1/description
http://www.geneontology.org/formats/oboInOwl#hasDefinition

Description Logic queries submitted to Aber-OWL available at http://aber-owl.net/queries.log. The query log contains the ontology that has been queried, the kind of query made (retrieving sub-, super- or equivalent classes), the number of classes returned, and the time it took to execute the query.

Future directions

In the future, we aim to further develop Aber-OWL in two major directions. First, we intend to explore how much of the semantics of ontologies can be made available in real time through an ontology portal. Currently, in Aber-OWL, we are using the ELK reasoner [13]. However, a large number of highly optimized reasoners are available, including some for more expressive fragments of OWL. We intend to evaluate some of these reasoners, based on the results achieved in the OWL Reasoner Evaluation challenges [2]. However, the theoretical limitations of non-tractable reasoning in OWL will remain a challenge, in particular with user-defined queries which may result in query times becoming too high. One solution to avoid this pitfall with more expressive fragments of OWL (or complete OWL 2) could be to set an upper limit for query answer time and fail if a query cannot be answered in that time, essentially resulting in incomplete reasoning. Nevertheless, such an approach could work if the majority of queries can be answered quickly.

Our second main aim for future development is to demonstrate additional functionality and novel types of bioinformatics applications that make use of inferences over ontologies. As our intended users fall in two categories (bioinformaticians/ontologists and domain experts), this step also takes two directions. For ontology developers in the biomedical domain, it is often difficult to evalu-

ate the consequences of a change made to an ontology, since the ontology may be imported in multiple other ontologies. For example, a single change in the Gene Ontology [1], which is imported by a large number of other ontologies, can have a significant impact on any of the other ontologies, such as resulting in incoherent class definitions or leading to inconsistency. At the moment, such consequences are not visible to the ontology developers. Aber-OWL has the potential of immediately showing the consequences of such a change across the range of ontologies it contains, essentially serving as a continuous integration environment for distributed development of ontologies.

Our other target audience, the domain experts, often work with ontologies as graph structures [9] that are used in visualization and data analysis. We intend to generate and visualize ontology graph structures, including the graph structures induced by axiom patterns [7], in addition to the subsumption hierarchy currently available through Aber-OWL.

Conclusion

Despite Aber-OWL being relatively new, we have already established a small user base, mainly for the REST API services. We have also demonstrated that reasoning even over a large set of ontologies is now a possibility and can be performed efficiently [15], and that novel kinds of applications can be developed which rely on automated reasoning and semantic query. These applications may even lead to new data- or text-mining methods that reveal new insights into a domain of knowledge [8]. In the future, we hope that Aber-OWL will establish itself as an ontology repository in the biological and biomedical domain that makes the semantics of ontologies and inferences over them available to a wide range of users.

References

1. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, M.J., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Tarver, L.I., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G.: Gene ontology: tool for the unification of biology. Nature Genetics 25(1), 25–29 (May 2000), http://dx.doi.org/10.1038/75556
2. Bail, S., Glimm, B., Jiménez-Ruiz, E., Matentzoglu, N., Parsia, B., Steigmiller, A. (eds.): ORE 2014: OWL Reasoner Evaluation Workshop. No. 1207 in CEUR Workshop Proceedings, CEUR-WS.org, Aachen, Germany (2014)
3. Belleau, F., Nolin, M., Tourigny, N., Rigault, P., Morissette, J.: Bio2RDF: Towards a mashup to build bioinformatics knowledge systems. Journal of Biomedical Informatics 41(5), 706–716 (Oct 2008), http://dx.doi.org/10.1016/j.jbi.2008.03.004
4. Bienvenu, M., ten Cate, B., Lutz, C., Wolter, F.: Ontology-based data access: a study through disjunctive datalog, csp, and mmsnp. In: PODS. pp. 213–224 (2013)
5. Cote, R., Jones, P., Apweiler, R., Hermjakob, H.: The ontology lookup service, a lightweight cross-platform tool for controlled vocabulary queries. BMC Bioinformatics 7(1), 97+ (2006), http://dx.doi.org/10.1186/1471-2105-7-97

6. Grau, B., Horrocks, I., Motik, B., Parsia, B., Patelschneider, P., Sattler, U.: OWL 2: The next step for OWL. Web Semantics: Science, Services and Agents on the World Wide Web 6(4), 309–322 (November 2008), http://dx.doi.org/10.1016/j.websem.2008.05.001
7. Hoehndorf, R., Oellrich, A., Dumontier, M., Kelso, J., Rebholz-Schuhmann, D., Herre, H.: Relations as patterns: Bridging the gap between OBO and OWL. BMC Bioinformatics 11(1), 441+ (2010)
8. Hoehndorf, R., Schofield, P.N., Gkoutos, G.V.: Analysis of the human diseasome using phenotype similarity between common, genetic, and infectious diseases. Scientific Reports 5, 10888 (2015)
9. Hoehndorf, R., Schofield, P.N., Gkoutos, G.V.: The role of ontologies in biological and biomedical research: a functional perspective. Briefings in Bioinformatics (March 2015), http://bib.oxfordjournals.org/content/early/2015/04/10/bib.bbv011.abstract
10. Hoehndorf, R., Slater, L., Schofield, P.N., Gkoutos, G.V.: Aber-OWL: a framework for ontology-based data access in biology. BMC Bioinformatics 16, 26 (2015), http://www.biomedcentral.com/1471-2105/16/26/abstract
11. Horridge, M., Drummond, N., Goodwin, J., Rector, A., Stevens, R., Wang, H.: The Manchester OWL Syntax. Proc. of the 2006 OWL Experiences and Directions Workshop (OWL-ED2006) (2006)
12. Horrocks, I.: OBO flat file format syntax and semantics and mapping to OWL Web Ontology Language. Tech. rep., University of Manchester (March 2007), http://www.cs.man.ac.uk/ horrocks/obo/
13. Kazakov, Y., Krötzsch, M., Simancik, F.: The incredible elk. Journal of Automated Reasoning 53(1), 1–61 (2014), http://dx.doi.org/10.1007/s10817-013-9296-3
14. Noy, N.F., Shah, N.H., Whetzel, P.L., Dai, B., Dorf, M., Griffith, N., Jonquet, C., Rubin, D.L., Storey, M.A.A., Chute, C.G., Musen, M.A.: Bioportal: ontologies and integrated data resources at the click of a mouse. Nucleic acids research 37(Web Server issue), W170–173 (July 2009), http://dx.doi.org/10.1093/nar/gkp440
15. Slater, L., Gkoutos, G., Schofield, P.N., Hoehndorf, R.: Using aber-owl for fast and scalable reasoning over bioportal ontologies. In: Proceedings of International Conference on Biomedical Ontologies (ICBO). pp. 72–76 (July 2015), http://icbo2015.fc.ul.pt/ICBO2015Proceedings.pdf
16. The Uniprot Consortium: The universal protein resource (uniprot). Nucleic Acids Res 35(Database issue) (January 2007), http://view.ncbi.nlm.nih.gov/pubmed/17142230
17. Usbeck, R., Ngonga Ngomo, A.C., Bühmann, L., Unger, C.: HAWK - hybrid Question Answering over linked data. In: 12th Extended Semantic Web Conference, Portoroz, Slovenia, 31st May - 4th June 2015 (2015)
18. Xiang, Z., Mungall, C.J., Ruttenberg, A., He, Y.: Ontobee: A linked data server and browser for ontology terms. In: Proceedings of International Conference on Biomedical Ontology. pp. 279–281 (2011)